

[75] Inventors: Jari Hamalainen, Tampere; Zhi Chun Honkasalo, Vantaa; Harri J. Hiisi, all of Finland

[73] Assignee: Nokia Mobile Phones Ltd., Salo, Finland

1
SYSTEM FOR TRANSMITTING PACKET
DATA IN DIGITAL CELLULAR TIME
DIVISION MULTIPLE ACCESS (TDMA) AIR
INTERFACE

BACKGROUND OF THE INVENTION

The present invention relates to the transmission of packet information in the air interface of a packet radio system. The general range of application is any digital cellular system based on TDMA, Time Division Multiple Access.

The majority of current cellular networks provide effective data and speech services based on circuit switched technology. However, the utilization of transmission resources in circuit switching is unoptimal, because the transmission connection is maintained throughout the contact irrespective of the fact whether information is transmitted or not at a given moment. Transmission resources are shared by multiple users, which means that the reservation of circuit switched connection for one subscriber only unnecessarily uses up transmission resources from other subscribers. The burstiness of data services also is a drawback in circuit switched systems. The utilization of the channel can actually be enhanced by applying packet switched information transmission. On the other hand, packet switching should be used only when an actual need arises, because the burstiness of data services is high and this may cause interference in circuit switching.

The future third generation cellular system UMTS (Universal Mobile Telecommunications System) must be able to transmit both circuit switched and packet data transmission, such as ISDN (Integrated Services Digital Network) and ATM (Asynchronous Transfer Mode) transmission. Now the key factor is the air interface, where an advanced multiple access technology is employed; by means of this, the channels supporting different types of services must be effectively multiplexed in the air interface both to and from the radio channel. The conference publication "Mobile and Personal Communications, 13-15 December 1993, Conference Publication No. 387, IEE 1993" includes the article "A Reservation Based Multiple Access Scheme for a Future Universal Mobile Telecommunications System" by J. M. DeVille, which describes the requirements to be set for the air interface of a UMTS system. For example, multiple access must be able to utilize the inactivity of the information source by granting a physical channel only when there is activity on the logical channel, and to support different bit rates so that time slots in the frame are allocated to the logical channel according to the needs of the situation.

In order to satisfy these and other requirements, there is suggested the multiple access control method PRMA++ (Packet Reservation Multiple Access), which is part of a design for third generation cellular systems related to the transmission of packetized speech and data. PRMA++ can thus be used as multiple access control both in packet switched and circuit switched transmission. The PRMA++ method concentrates on using one time slot in the transmission of packet data.

On the radio channel, PRMA++ uses Time Division Multiple Access TDMA. This allows the subscriber to share the transmission resources of the radio channel. The TDMA frame is divided into time slots, where the transmitted burst carries the data as well as signals connected to channel coding, notifications etc. In the uplink direction, which is the direction from the mobile station to the network (base station:), there are two types of time slots: reservation or R-slots, where only channel request bursts are transmitted.

and information transmission or I-slots, which are only used for transmitting information bursts. In the channel request burst, the mobile station uses an Air-Interface Channel Identifier containing the network address of the mobile station, which address identifies the logical channel, and where it requests one or more time slots from the frame, according to the needs of the moment. In the downlink direction, i.e. from the network (base station) to the mobile station, there are likewise two types of time slots: I-slots reserved for transmitting information, and acknowledgement or A-slots. When the mobile station requests access to the network, the base station acknowledges the request on A-slots by transmitting the address of the subscriber and the number of the I-slot. From this onward, the said I-slot is reserved for the use of the mobile station.

Let us suppose that the number N of the PRMA++ time slots in one TDMA frame is a system configuration parameter. On the uplink channel, one TDMA frame now contains one R-slot and $N-1$ numbers of I-slots. All mobile stations start transmission by transmitting a channel request on the R-slot, and if several mobile stations use the same R-slot for transmitting the request, collisions may occur. The downlink TDMA frame includes, in addition to the above mentioned A-slot for acknowledging channel requests transmitted on the R-slot and I-slots, also a fast FP (Fast Paging) slot constituting the fast paging channel, on which the mobile station is notified of incoming data transmission and of information transmission slots.

The mobile station starts transmission by channel request on the uplink channel on an R-slot, which is used for this purpose by all mobile stations of the same cell. The base station acknowledges the received channel request on the acknowledgement burst on the downlink A-slot. If no requests are transmitted on the R-slot, or if on the channel there are collisions, identified by the base station, the base station transmits an idle flag on the acknowledgement burst of the respective A-slot, so that the mobile station understands to repeat the channel request after some time. In case the channel request sent on the R-slot was correctly received, but there are no free time slots for the transmission, the mobile station is notified of this on the next downlink time slot. The mobile station queues for access until a free time slot is found.

The R-slot contains a training sequence, address of the mobile station, number of requested information slots and a circuit switched flag. The flag informs whether the reservation is valid for the duration of the packet or longer. The channel is reserved, until an order for cancelling the reservation arrives. The A-slot acknowledgement burst contains the address of the requesting mobile station as well as the channels that are granted for traffic. The mobile station receives the acknowledgement burst, whereafter it tunes the receiver and transmitter to the allocated channel. Traffic on this channel is started, and it continues as long as there is data or speech to be transmitted. In packet data transmission, the number of bursts—bursts here meaning packets that are transmitted after one channel request—can be constant.

The base station uses the fast paging slot, FP-slot, to notify the mobile station of an incoming packet. The mobile station listens to the FP-channel and decodes all received messages in order to notice its own identifier. The time slot on the fast paging channel contains a list of those I-slots that are allocated for the mobile station. The mobile station acknowledges its own paging by transmitting an acknowledgement in the FP-acknowledgement slot.

According to what was said above, it is characteristic of the suggested UMTS system both in the uplink and in the

0025325 "022399
66E220" 2255260

downlink directions that physical channels are not allocated for connections which are not active at a given moment, and hence they do not reserve capacity in vain. The channels are always reserved by the same protocol, both in the case of circuit switched and packet transmission. The allocation of the channels is not dynamic, wherefore the channels reserved for packet usage cannot easily be altered. Reservation, fast paging and acknowledgement slots are given slots, and the state of the art does not comment on altering these. Moreover, the known method does not pay particular attention to the symmetry or asymmetry of packet transmission when creating a transmission channel.

SUMMARY OF THE INVENTION

The present invention relates to a packet data transmission system in an air interface, the said system having eliminated the above described drawbacks. In accordance with the invention, there is created a flexible system whereby channels can be flexibly created and altered according to the situation in hand, which enables an extremely effective utilization of the channel resources and provides possibilities for using different data rates. Here the system is called Variable Rate Reservation Access VRRRA.

Aspects of the invention are defined in the appended claims.

Channels are allocated dynamically so that a variable number of time slots in the cell is reserved for packet usage, and the rest of the time slots are used for circuit switched services, including speech. The mobile station can select the number of employed time slots, and the network adjusts to that, so that even a simple one-slot mobile station can use the packet services. In case several time slots are reserved for the mobile station, each of which slots constitutes a sub-channel, for each slot there is designed error correction, interleaving and a corresponding frame length. If several time slots are required for one mobile station, there are reserved several of the said sub-channels, and each sub-channel uses the same error correction and interleaving algorithm. Thus there is needed only one algorithm. In the information transmission of the subscriber, the MAC layer at the transmitting end distributes the subscriber data to be carried through several sub-channels, and the MAC layer at the receiving end receives the frames of the sub-channels and compiles them to a complete subscriber data. From the point of view of the base station, each time slot is thus similar on the physical level. Now one and the same mobile station can use for instance two time slots, or one mobile station can use one and another can use the other. The prior art applies algorithms for units of one, two, three etc. time slots, in which case the base station respectively deals with channels comprised of one, two, three etc. time slots. A data stream conducted to the radio channel through the radio interface is multiplexed into several "pipes", i.e. into said independent sub-channels, and after receiving the packets, the data is again demultiplexed from the "pipes" to a data stream.

BRIEF DESCRIPTION OF THE DRAWINGS

An exemplary embodiment of the invention is further described with reference to the appended drawings, where:

FIG. 1 depicts the frame structure of a prior art system;

FIG. 2 is an example of using all time slots of the TDMA frame;

FIG. 3A depicts the principle when all time slots are not used;

09255325.022349
662220.5225260

FIG. 3B is an example of using one single time slot of the TDMA frame;

FIG. 4A depicts a fast paging/acknowledgement burst;

FIG. 4B depicts a modified access burst;

FIG. 5 depicts a case without any determined R-slots;

FIG. 6 depicts some structures of the control fields;

FIG. 7A depicts an asymmetric mobile terminated packet transmission;

FIG. 7B depicts an asymmetric mobile originated packet transmission;

FIG. 8 depicts the fields of the acknowledgement burst;

FIG. 9 depicts an alternative to asymmetric transmission;

FIG. 10 depicts a symmetric packet transmission;

FIG. 11 depicts the access server queue principle; and

FIG. 12 depicts a block diagram of a conventional implementation of logical channels in a mobile station and in a base station.

DESCRIPTION OF THE PREFERRED EMBODIMENT

In the examples below, it is assumed that the frame structure in the network is such that one TDMA frame consists of eight time slots, in the same fashion as in the known GSM system, but it is understood that the number of time slots is a system configuration of free choice. The number of those time slots that are allocated for packet radio can depend on the use demand for packet data in the cell. If there are only a few packet service subscribers, it is sufficient to allocate just a few time slots in the frame for packet data, and if there are several subscribers, all eight time slots are reserved. It is up to the operator to configure how many time slots must be reserved.

FIG. 2 illustrates an extreme case where all eight time slots in the frame are reserved for transmitting information data. In a logical channel arrangement in the downlink, the fast paging FP-slots and the acknowledgement A-slots for the requests transmitted by the mobile stations logically share the control time slot, which is the first slot in each frame. In the uplink, the random access slot R is repeated at standard frequency. Every second R-slot can be replaced by a fast paging acknowledgement slot, FPA-slot. This can be used for acknowledging reception of fast paging, in case other acknowledgements are not in use. In this example the control channels both on the downlink and uplink channels are thus formed of every eighth time slot. The rest of the time slots, enumerated from 1 to 7, are used for transmitting packet data. In this extreme case, the data transmission rate is maximal, because the whole frame is in use, i.e. transmission takes place on every time slot.

FIGS. 3A and 3B illustrate an extreme case where only one time slot in the frame is reserved for transmitting packet data. Now the seven remaining time slots are free to be used for instance for circuit switched purposes. FIG. 3A shows in principle how the logical channel is formed. According to the drawing, the first time slot in each frame is reserved for transmitting packet data, and the slots 2 . . . 8 are free for other use. The time slots reserved for transmitting packet data are illustrated as darkened squares. The consecutive first time slots of each frame thus form a channel, their "frame" consisting of one time slot. FIG. 3A illustrates the logical channel created at the tip of the arrows. In this extreme case, the transmission bursts take place on every eighth time slot with respect to the TDMA frame, wherefore the data transmission rate is low.

FIG. 3B depicts some uplink and downlink channels formed according to the principle illustrated in FIG. 3A. They contain FP- or FP/A-slots and R-slots, as well as I-slots for information data. In the uplink direction, the R-slots may have permanent locations, in the drawing every fifth slot, which means that an R-burst can be transmitted on every fifth TDMA frame of the network, on the first time slot thereof. In the downlink, these permanent locations can be provided with fast paging and acknowledgement slots, so that for instance every tenth frame can be a fast paging frame FP, and every tenth frame again an acknowledgement frame A. As in the drawing, fast paging and acknowledgement can also be combined to a fast paging and acknowledgement slot FP/A. Depending on the maximum length of the reservation time, the frequency of the control channels both in the uplink and in the downlink can be lower, for instance every ninth, every thirteenth etc. time slot, instead of every fifth as in FIG. 3.

FIGS. 2 and 3B depict two extreme cases for allocating the time slots. In between these cases there remain those where 2, 3, 4, 5, 6 or 7 time slots are allocated for packet data. The principle is evident on the basis of what is explained above, and these cases are not dealt with in more detail here. The number of the allocated time slots determines the transmission rate of packet data, which means that the required transmission rate can serve as one criteria for allocating the time slots.

The logical channel structure may be different in different cases where the number of time slots reserved for packet data also differs. If two times slots are reserved for packet data, one of them can be used for data only and the other for control (FP, A, R). Another possibility is to use one for data only, and the other for both control and data, because there is not much need for control with only two time slots in use. In this case a combined control/data slot has a smaller data capacity. When all I-channels are reserved for transmitting the user's packets, control is not needed any more. Now the control slot or the logical control channel can be granted for transmitting information data, i.e. more I-capacity is obtained. As soon as an I-channel becomes free, a new logical control channel must be created.

In FIG. 3B the control slot is repeated as every fifth time slot, but the repeating rate can also be other than that. The structure of the FP- and A-bursts can also depend on the number of allocated time slots, and they can also be combined, in the above described fashion, so that they share the same time slot, as is illustrated in FIG. 3 B. Now for instance every second control slot is paging and every second acknowledgement.

If the Invention is applied for example to the GSM or PCN systems, the suitable fields for paging and acknowledgement bursts would be such as are illustrated in FIG. 4 A. The fast paging FP needs the temporary subscriber identity of the mobile station MS, and the acknowledgement A needs a random number used on the R-slot as well as a timing advance TA. In the request, the mobile station transmits a random number, which the base station retransmits in its acknowledgement A in order to notify the mobile station that the transmitted message was received.

In the drawing of FIG. 3B, the R-slots in the uplink are permanent. This is not, however, compulsory. FIG. 5 illustrates an arrangement where the mobile station is not allocated a given R-slot for the access burst R in the uplink. In the downlink, there may be logically allocated connection-wise time slots for fast paging. The base station may transmit fast paging FP at regular intervals, and the paging may also

contain an identifier of occupied uplink channels, such as a channel bitmap, which notifies which channels are free or occupied in the uplink. Therefore the control system of the radio resources of the base station must follow the used time slots.

According to FIG. 5, the base station transmits fast paging FP on one time slot of the TDMA frame. This is represented by the upper frame, which illustrates the reception of the mobile station MS. The mobile station acknowledges the paging on the respective time slot in the uplink. This is represented by the lower frame, which illustrates the frame of the mobile station MS in the uplink. However, the mobile station does not have a given R-slot for transmitting the access burst, but it may be transmitted on any one or several of the free time slots of the TDMA frame in the uplink. The base station acknowledges access on a downlink slot connected to this uplink slot. If the mobile station transmits an access burst for instance on the third time slot, the base station acknowledges it on the following third time slot of its frame. The acknowledgement contains a reference to the allocated time slot. In case the corresponding downlink time slot required for acknowledgement is not free, but it is reserved for data transmission on another channel, the control burst simply steals this time slot for its own use. This means that if there is a need to transmit control data, and the required time slot is dedicated for transmitting information (I-slot), the problem is solved so that instead of the information of another channel, there is simply transmitted the control data, such as the said acknowledgement burst, provided with a steal flag. From the steal flag the mobile station subjected to stealing knows that there is a stolen time slot on which acknowledgement is transmitted to some other mobile station, and the expected information is not coming until the next time slot.

FIG. 4A depicted the fields of the fast paging/ acknowledgement burst. FIG. 4B illustrates an access burst to be transmitted on an R-slot. The structure of a GSM access slot as such is not suited to be used on the R-slot in the system of the present invention. The said access burst is presented in FIG. 4B. It contains a 41-bit synchronizing part and 36 encoded information bits, which result from an error correction algorithm for encoding a databit of 8 users. Eight databits are insufficient for the purposes of the present invention, wherefore their number must be increased. This is solved by applying a new encoding ratio. Accordingly, 12 databits and 6 CRC bits are encoded by applying $\frac{1}{2}$ FEC (Forward Error Correction), resulting in 36 encoded bits. Thus at least 12 bits can be obtained for employed data instead of the 8 bits of GSM. Therefore the channel encoding must be changed. As an alternative, it is possible to apply the channel encoding method of the GSM access burst, so that only 8 databits are available. Now we have obtained 12 databits on an R-slot, and the databit number of the access burst has reached the aim, because the sum of the field of random reference number, with a length of 9 bits, and of the field giving the number of time slots, with a length of 3 bits, is exactly 12 bits. Alternatively, if the number of the bits informing the random reference is increased (in GSM the number is 5 bits), the data section of the same 12 bits can be used for informing some priority bits, too. These alternatives are illustrated in FIG. 4B. Random reference is applied in the same fashion as in the current GSM system, so that in its access burst the mobile station transmits a random number to the base station, which in its acknowledgement retransmits the same number plus channel information. Now the mobile station in question identifies the acknowledgement to be intended for it. FIG. 12 depicts a block diagram of a

conventional implementation of logical channels in a mobile station and in a base station. -At page 11, after line 12, insert the following paragraphs:

FIG. 12 is a block diagram of a conventional, prior art implementation of logical channels in a mobile station and in a base station. This conventional implementation is suitable for practicing the instant invention. On the transmitter side the error coding is performed in a channel coding block 12. If the input data is a speech signal, it is first coded in a speech coding block 10. The signal is then interleaved in an interleaving block 14, encrypted in an encryption block 16 and modulated to the carrier frequency in a modulation block 18. The modulated signal is then transmitted to the radio channel 20. The above-mentioned stealing bit is also set in the interleaving block 14, as controlled by a control block 22.

In the receiver side the carrier frequency signal is received from the radio channel 20 and is demodulated to the base-band frequency in a demodulation block 24. The signal is then decrypted in decryption block 26 and de-interleaved in a de-interleaving block 28, which also yields the stealing bit information for the control block 22. The error decoding is performed in a channel decoding block 30, and if the received data is speech, speech decoding is accomplished in a speech decoding block 32. All the mentioned transmitter and receiver blocks are controlled by the control block 22.

A more detailed description of the functions of the transmitter and receiver blocks can be found in, e.g., a publication by Michel Mouly and Marie-Bernadette Pautet; The GSM System for Mobile Communications, 1992, France.

FIG. 6 illustrates a possible structure suggested for the fields of the R, A and FP slots. They are not GSM-specific. The acknowledgement of the access burst, constituting a random number and the number of desired time slots, is transmitted on an A-slot. The structure of the A-slot contains a random reference, a bitmap of the reserved channels and a time advance, calculated by the base station. Moreover, it may contain a reference to those time slots that were transmitted on the R-slot. This prevents the same channel from being granted to multiple subscribers. The fields of the fast paging contain the temporary identity of the mobile station and the bitmap of allocated time slots. The temporary identity of the mobile station is an identity used for the MS in the packet transmission mode.

In the above specification we have explained channel formation in general, but we have not paid attention to the direction of data packets. User information, i.e. data packets, are transmitted on I-slots reserved for the transmission of data packets by using normal bursts. Naturally a radio channel can be reserved symmetrically, by reserving an equal number of time slots in both directions. Generally data transmission is, however, asymmetric, and symmetric reservation means that resources are wasted in one of the directions.

An asymmetric transmission can be realized in two different ways. In the first alternative the MAC (Media Access Control) protocol is half duplex. The Information slots, I-slots, are reserved only in one direction at a time. MAC first investigates which direction the packet is going to be transmitted to, and reserves either an uplink or a downlink channel, according to the required direction. Acknowledgement is not used with information slots on the MAC level. Acknowledgements are carried by the link layer protocol, and for acknowledgement transmission there is reserved an information slot in similar fashion as for the user's data.

FIG. 7A illustrates the time slot structure in an asymmetric mobile terminated transmission. In the downlink the base

10

15

30

55

65

receives the acknowledgement, step f, and transmits again on the slots 1-6 of the next frame. This is repeated as long as there is data to be transmitted. In the drawing, the darker squares represent those slots that are reserved for one connection only. The asymmetry is thus clearly visible. It is also pointed out that in between successive bursts, there can be applied frequency hopping, i.e. the consecutive slots of a frame use different physical channels.

FIG. 10 illustrates a symmetric transmission. Data is transmitted in one direction, and the corresponding slot of the opposite direction is used for acknowledgements. In the drawing the mobile station requests a channel on the R-slot, which request is acknowledged by the base station in the downlink acknowledgement burst on an A-slot. The mobile station is allocated one slot, which is the first information slot I. On this slot, it transmits a packet burst, which is acknowledged by the base station on the first corresponding slot. Thus the transmission and acknowledgement alternate on corresponding slots in the uplink and downlink directions. Other slots are free or used by other mobile stations, and part may be reserved for circuit switched services.

The mobile station MS and the packet arrangement in the network may exchange various parameters at the beginning of the packet session. This is useful because thus the mobile station informs the network as to the number of the slots in the TDMA frame that it can handle during data transmission. The mobile station can be so simple in structure that it is capable of dealing with only one slot, whereas the network can handle all slots of the frame. When the mobile station informs the network of this feature at the beginning of the session, the network immediately knows to allocate only one slot for the mobile station. At the beginning the network also informs the mobile station as to the packet slots in the network, so that the mobile station finds out what kind of logical channel structure the cell has.

Irrespective of the form of the access burst, the system allows for a priority value to be included in the reservation request on the R-slot. There can be several different priority levels with different binary values. In the base station, the queue system may observe the priority of the data transmission requests. There is a maximum time value, for the duration when random access is valid. This prevents an incorrect use of the timing advance. When the base station receives the R-slot, it arranges the received requests in a queue. The channels reserved in the requests are allocated either in the FCFS order, or based on the prioritised FCFS, if the request includes a priority value. The base station must include a time stamp for each received R-slot in order to keep track of the requests for which the maximum timing has been exceeded. Timing reserves a given duration for the request to be valid, and if the timer elapses before the request is fulfilled, it is removed from the queue.

FIG. 11 illustrates the queue principle. When the access server receives an R-slot, it inserts it in the queue. The server selects the request having next turn, i.e. being oldest in duration or highest in priority, and finds out whether a requested number of slots is available. If there are enough free slots, the acknowledgement is transmitted on an A-slot.

If there are not as many free slots as was requested, there are two possibilities: 1) the request is queued until a sufficient amount of free capacity is found, or 2) the mobile station is given as many slots as there are free. In the first case, a message must be sent to the mobile station on the acknowledgement slot. This prevents unnecessary timer expiration. For those R-slots for which the time has elapsed, a retransmission algorithm is used.

The present invention can be applied to any digital TDMA cellular system. It is not necessary to alter the burst structure of the systems where the invention is applied, for instance GSM and PCN systems. The structure of logical channels renders several different possibilities and enables a dynamic allocation of resources and a flexible access protocol.

When applying the system of the invention to a GSM system, there are three different modes available:

1. Idle: the mobile station does not use packet data services, but only circuit switched services. Now it works as the current GSM phone.

2. Packet half active: the mobile station is in virtual connection state for packet services (no physical channel), but is currently not in the receiving or transmitting data state. It is listening to the standard GSM paging channel to receive data packets. In this mode, only the paging channel is listened to, which saves the batteries because listening takes place fairly seldom. When a packet is coming in, there is transmitted an ordinary paging, the reason code being "arrival of packet". Now the mobile station shifts to mode 3, where it listens to the FP channel and can receive the packet.

3. Packet active: the mobile station is transmitting or receiving packet data. When not actively transmitting a packet, it listens to the FP channel to receive packets, as well as to the standard GSM paging to receive speech. This is called the FP DRX state (Fast Paging Discontinuous Reception).

When the transmission of packets in mode 3 is interrupted for some reason, the timer is started. When it elapses to a predetermined value, the mobile station is shifted from mode 3 to mode 2, so that the physical channel becomes free. When the mobile station requests to initiate virtual connection, it exchanges parameters with the network, such as exchange of encryption keys, initiation of encryption, identification etc. At the beginning of the virtual connection, the mobile station informs the network of the number of time slots that it is made for. Thus the network knows not to transmit data on eight slots, if the mobile station is made for one slot only.